Advanced Performance Modeling with Combined Passive and Active Monitoring

Alex Sim, CRD, LBNL
Constantine Dovrolis, Georgia Tech

New project

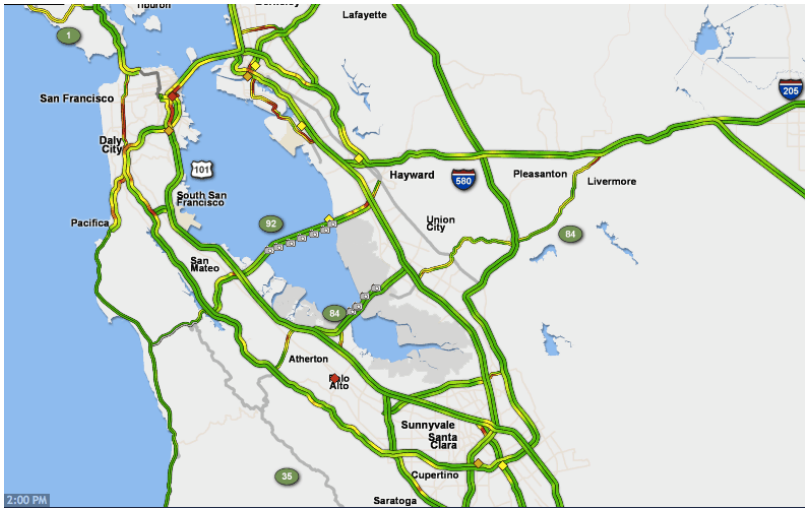
- **Terabit networking for extreme scale science project**
  - **Started 10/2011 at LBNL**
  - **Started 1/2012 at GT**
  - **Explore fundamental questions on the relationship between monitoring and estimation of network resource performance**

Overview

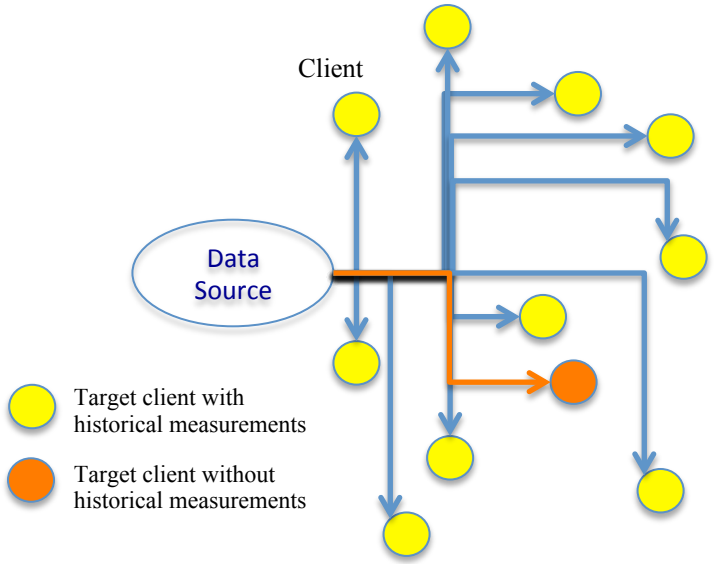
- **Objectives**
  - Investigate and develop network performance estimation model
    - Based on historical, passive and active performance measurements
  - Develop a general purpose, efficient and reusable performance estimation framework.
- **The predictive performance estimation model**
  - helpful in optimizing the performance and utilization of fast networks for scientific collaborations.
    - For scheduling data transfers and sharing resources with predictable performance
    - For remote data access with probabilistic variability estimate
  - estimate future network usage and the latency.

Challenges

- **Challenges**
  - Access to historical monitoring measurement collection
  - Access to current monitoring measurements
  - Hybrid performance estimation model
  - Long-term performance estimation
  - Performance variability and uncertainty estimation



- **Example**
  - Highway traffic pattern giving a time estimation for travel planning (e.g. roughly 1.5 hours from Berkeley to SFO on Monday morning 8:30am, or ~40 minutes around 1pm).



Measurements

- Historical network monitoring logs for analyzing network traffic pattern on shared network resources
- Passive measurements from the data movement tools
- Active measurements from probing systems

Technical background

- **Two previous work**
  - **Predictability of large transfer TCP throughput**
- **SIGCOMM 2005, COMNET 2007**
  - **He, Q., Dovrolis, C., and Ammar, M., "On the Predictability of Large Transfer TCP Throughput," In the Computer Networks (COMNET) Journal, 51(14), pp. 3959-3977, 2007.**
  - **Formula-Based (FB) prediction**
  - **History-Based (HB) prediction**
- **TPDS 2011**
  - **Kim, J., Chandra, A., and Weissman, J. B., "Passive Network Performance Estimation for Large-scale, Data-intensive Computing," IEEE Transactions on Parallel and Distributed Systems (TPDS), Vol. 22, Issue 8, p. 1365-1373, Aug. 2011.**
  - **Estimation model in the absence of any prior path-specific information**

Project tasks

- **studying and exploring performance estimation models and statistical prediction**
  - investigation of the hybrid estimation models and algorithms for end-to-end performance estimation
  - estimation with confidence intervals
  - performance prediction for future time windows
  - variability and uncertainty estimation/prediction
  - designing the end-to-end measurement collection
  - coordinate with perfSONAR and other projects to collaborate on active measurement data
  - prototype of Data Mover using the estimation
- **validation and experiment**
  - conduct a study with collaborating scientific communities to validate the estimation models in fast network environment as well as traditional shared network environments

Georgia Tech

Current Status

Questions

- LBNL: Alex Sim <asim@lbl.gov>
- Georgia Tech: Constantine Dovrolis <dovrolis@cc.gatech.edu>